MULTIRESOLUTION CODING OF IMAGE AND VIDEO SIGNALS

Bernd Girod *Frank Hartung* *Uwe Horn**
Telecommunications Institute I,
University of Erlangen-Nuremberg,
D-91058 Erlangen, Germany
{girod,hartung}@nt.e-technik.uni-erlangen.de Uwe.Horn@gmd.de

Invited paper

ABSTRACT

Multiresolution image and video coding schemes offer both excellent coding efficiency and the ability to support scalability. This paper gives an introduction to the principles of multiresolution coding of image and video signals. Besides critically sampled subband pyramids, oversampled pyramid decompositions are discussed. Oversampled subband pyramids are often better suited for scalable coding schemes. Solutions to the bit allocation problem for subband coders with and without quantization noise feedback are presented. Finally, we briefly review spatio-temporal pyramids as the most promising approach to scalable video coding.

1 INTRODUCTION

The integration of images and video into multimedia systems requires efficient compression techniques of low complexity. Compression can take advantage of the statistical dependencies in images and video, and it should also exploit the limitations of human visual perception and omit irrelevant signal components. Multiresolution coding is not only able to address both principles in an elegant and efficient way, but can as well be utilized for scalable coding [1, 2, 3]. Scalability means that an already compressed signal can be decoded at different quality levels, using only an appropriate subset of bits. The reconstruction of lower quality renderings not only requires a lot fewer bits, but also significantly less computation than the reconstruction of the full quality reconstructed image.

This paper gives an introduction to multiresolution image and video coding. We motivate why subband coding and especially subband pyramid coding are efficient for images and video. Besides critically sampled subband pyramids, oversampled pyramid decompositions are discussed as well since they are better suited for applications requiring scalability. We also review recent analytical solutions to the bit allocation problem for multiresolution coders with and without quantization noise feedback. Finally, we briefly discuss spatio-temporal pyramids as the most promising approach to scalable video coding.

*now with Institute for Media Communication, German National Research Center for Information Technology (GMD), Schloss Birlinghoven, D-53754 St. Augustin, Germany.

A more detailed introduction to multiresolution decomposition and subband image coding can be found in [4, 5, 6, 7]. A good overview about multiresolution video coding is given in [8, 9].

2 MULTIRESOLUTION IMAGE CODING

Multiresolution image coding is based on the analysis of an image at a hierarchy of resolutions or scales. It is motivated by the observation that typical images possess a power spectral density that falls off rapidly towards high spatial frequencies. The rate of decay is most rapid at low frequencies and smaller at high frequencies. Thus, most of the signal energy is contained in lower frequency components, and a coarsely sampled version capturing only the lowpass components is already a good approximation in the mean squared error (MSE) sense. Additionally, the human visual system is less sensitive to errors at high spatial frequencies. Rate distortion theory suggests that an optimal coder for stationary Gaussian signals of arbitrary power spectral density splits the original signal into spectral components of infinitesimal bandwidth and encodes them independently [10]. This is the basic motivation for subband coding. Of course, arbitrarily small bandwidths cannot be achieved for finite-size images, and they would not even be desirable in practical applications, since spatial localization of the basis functions representing the image would be lost entirely. Poor localization of image features such as edges after filtering tends to introduce subjectively annoying artifacts when combined with lossy compression of the subband signals.

2.1 Subband pyramid decompositions

Subband pyramid decompositions are a good compromise between frequency selectivity and spatial localization. Because of the typical shape of the power spectrum of natural images, pyramid decompositions produce subband signals that tend to be spectrally flat, such that simple memoryless coding of the individual samples within each band is justified and indeed often used in practice. Pyramid decomposition can be accomplished efficiently by cascading horizontal and vertical two-band filter banks as shown in Fig. 1. Each filtered signal is subsampled by a factor of 2

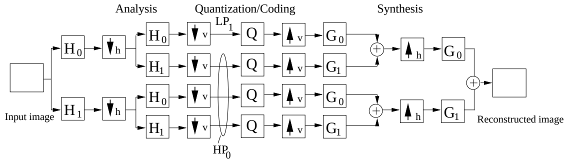

Figure 1: Commonly used filterbank structure to obtain critically sampled subband pyramid decompositions. H_0 and H_1 are lowpass and highpass analysis filters, G_0 and G_1 the corresponding synthesis filters. By recursively applying the analysis stage to the lowpass signal LP_1 additional layers are obtained. The three subbands marked HP_0 together contain the highpass signal components.

horizontally or vertically, such that the number of samples before and after subband decomposition is the same. The subband signals are quantized and encoded using appropriate fixed or, better, variable length coders. At the decoder, the subband signals are decoded, up-sampled and passed through a bank of synthesis filters. The output signals of the synthesis filters are finally added to yield the reconstructed signal.

The problem of designing analysis and synthesis filter banks such that the synthesized image is a perfect reconstruction of the original when the bit-rates are high is well-understood [4, 7] and need not be discussed here. Note that critically sampled subband pyramid decompositions contain the popular "Discrete Wavelet Transform" as a special case. Figure 2 shows an example of a wavelet decomposed image. While current image and video coding standards are based on the Discrete Cosine Transform (DCT), the discrete wavelet transform is very seriously considered as part of the JPEG 2000 standardization effort.

Many state-of-the-art subband pyramid image coders obtain increased coding efficiency by exploiting statistical dependencies between subbands. The basic scheme was introduced by Shapiro as the embedded zero-tree wavelet algorithm (EZW) [11]. The algorithm observes that if a coefficient at a low frequency is zero, it is highly likely that all the coefficients at the same spatial location at all higher frequencies will also be zero. Refinements of this idea have been presented, e.g., in [12].

With subband pyramid coding we can achieve scalability by extracting a subset of the subbands starting with the low frequency subbands. Note that reconstruction based on only a subset of subbands can be problematic in terms of perceptual quality, especially for finely tuned, long filters. Especially, aliasing cancellation in perfect reconstruction filter banks only works if all subbands are used for reconstruction. Therefore, critically sampled subband decompositions are less suitable for applications requiring frequent rendering of the image at reduced resolutions or quality levels [1, 13].

2.2 Oversampled pyramid decompositions

Besides critically sampled subband pyramids, oversampled pyramid decompositions can be employed as well. They were first introduced as the Laplacian pyramid by Burt

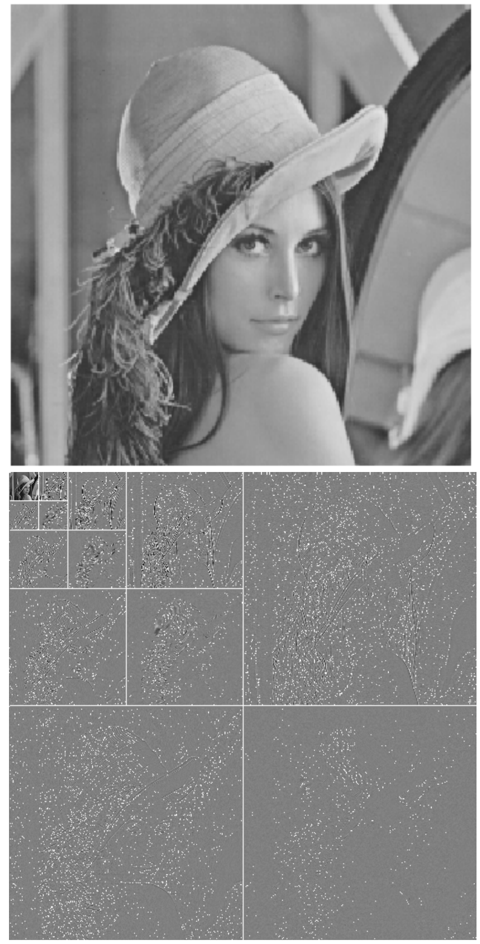

Figure 2: Original test image Lenna (top), and the same image decomposed into 13 bands by a discrete wavelet transform (bottom).

and Adelson in 1983 [14] and successively refined, e.g. by [8, 15, 3]. An input picture is first lowpass filtered and downsampled. By interpolating the low resolution image back to its original resolution a prediction of the original image is obtained. The prediction error can be viewed as a highpass subband signal which is not subsampled, unlike in critically sampled decompositions. The decomposition into a lower resolution image (LP) and a highpass error signal (HP) can be repeated for the lowpass image (LP) to generate additional pyramid layers. Note that the number of samples increases by up to 1/3 for 2:1 two-dimensional subsampling. On the other hand, one gains complete freedom in choosing appropriate filters for downsampling and interpolation. Lower resolution images within predictive resolution pyramids often have better subjective image quality than those obtained from critically sampled subband pyramids [8].

Interpolation to predict the higher resolution image can be based on either the filtered and subsampled version of the original image (open-loop coding) or its quantized version (closed-loop coding) (Fig. 3). The decoder is identical in both cases, hence open-loop and closed-loop coding can be freely combined. Burt and Adelson's Laplacian pyramid was introduced as an open-loop coder, and most researchers have considered this simpler scheme since. Closed-loop pyramid coding, however, which feeds back the quantization error in the spirit of Differential Pulse Code Modulation (DPCM), possesses several advantages over open-loop coding, as will be discussed in the next section.

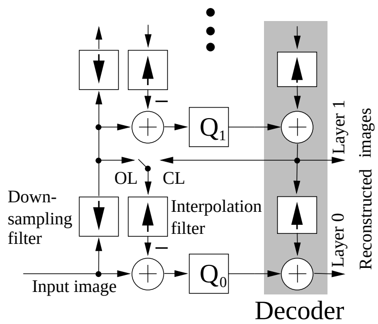

Figure 3: The first two stages of a pyramid codec for oversampled decompositions. OL and CL denote switch positions to select between open-loop and closed-loop coding. In contrast to open-loop coding, closed-loop coding employs quantization noise feedback at the encoder. The decoder is the same for both cases. LP_i denotes the lowpass signal, HP_i the highpass signal *within* layer i.

2.3 Bit Allocation

After an appropriate multiresolution decomposition, compression is commonly achieved by applying scalar quantization or vector quantization followed by entropy coding [16]. A detailed discussion of appropriate techniques can be found in [10].

A general problem in multiresolution coding schemes is the problem of optimal bit allocation. Since more than one quantizer is involved in the coding process (see Figs. 1 and 3) the question arises how to distribute an available maximum number of bits among the quantizers in an optimal way. 'Optimal' often means minimum MSE in the full resolution reconstructed image but other optimization criteria are possible as well [17].

For a given coding scheme, the bit allocation problem can always be solved experimentally by enumerating all possible bit allocations and selecting the optimal one which fulfills the bitrate constraint. This brute force approach is impractical for real-world applications. Under the assumption of a Gaussian source and high rate coding it is possible to derive analytical solutions for critically sampled subband coding schemes [5]. In the following, we review a generalization applicable to critically sampled as well as oversampled decompositions with and without noise feedback [18].

Let us consider the open-loop case first. We assume a stationary Gaussian input signal with memory and independent high-rate memoryless coding of each subband signal. Let R denote the available number of bits for which the MSE of the reconstructed image should be minimized, and let α denote the power transfer factor of the interpolation filter used in the synthesis stage. g_l and σ_l^2 specify the spectral flatness and variance of subband signal l. By using a Lagrangian approach we obtain the optimum rates r_l of a decomposition into L subbands as

$$r_l = \frac{R}{M} + \frac{1}{2}\log_2 \frac{\alpha^l \cdot g_l \cdot \sigma_l^2 / n_l}{\prod_{k=0}^{L-1} (\alpha^k \cdot g_k \cdot \sigma_k^2 / n_k)^{\frac{n_k}{M}}}, \qquad (1)$$

where $n_l = N_l / N_0$ is the ratio between the number of samples in layer l (N_l) and the number of samples in the full resolution layer (N_0). The redundancy due to oversampling is therefore $M = \sum_{l=0}^{L-1} n_l$. For critically sampled decompositions $M = 1$, for oversampled decompositions $M > 1$.

To obtain a solution for closed-loop pyramid coding schemes we have to take into account that, due to noise feedback, filtered quantization noise introduced in the previous stage is added to the subband signal *before* it is encoded. If we neglect that g_l depends on $r_{l+1}, \ldots, r_{L-1}$, we arrive at an analytical solution to the bit allocation problem as:

$$r_l = \begin{cases} \frac{1}{2}\log_2\left(\frac{\alpha \cdot g_l \cdot \sigma_l^2}{g_{l-1} \cdot \sigma_{l-1}^2} \cdot w_l\right), & l > 0 \\ R - \sum_{l=1}^{L-1} n_l \cdot r_l, & l = 0 \end{cases} \qquad (2)$$

$$w_l = \begin{cases} \frac{n_{l-1} - n_l}{n_l - n_{l+1}}, & \frac{n_{l-1} - n_l}{n_l - n_{l+1}} > 0 \\ 1, & n_l = n_{l-1} = n_{l+1} \end{cases}$$

with $n_L = 0$. The rather tedious derivation of (1) and (2) is due to Horn and Wiegand and can be found in [19]. Recently an equivalent solution for the special case of a regular decomposition with $n_l = \frac{1}{4^l}$ has been published independently of our work in [20]. In contrast to subband codecs without noise-feedback, the rates r_l for $l > 0$ are independent of the overall bitrate R and determined only by the statistics of the subbands. In layer 0 the remaining bits after quantization of all other layers are spent. A major drawback of the optimal solution for closed-loop coding is, that the distortion *within* lower resolution layers can become quite high compared to the distortion in the full resolution layer which is undesirable for scalable coding applications. A good heuristic is to use an optimal *open-loop* bit allocation for closed-loop coding [18].

Another interesting observation is that coding *with* noise feedback outperforms coding *without* noise feedback at higher bit rates. This can be derived from the solutions given in (1) and (2) and has been verified in coding simulations [18].

3 MULTIRESOLUTION VIDEO CODING

Several authors have tried to extend the idea of image subband coding to video coding by employing critically sampled 3-D subband coding schemes. Due to the associated delay and low coding efficiency, those approaches are not able to compete with motion compensated coding methods. It seems difficult to naturally include motion compensation within a critically sampled subband coding scheme [21, 22, 23, 24, 25].

Spatio-temporal resolution pyramids first proposed in [13] are a more suitable approach for scalable video coding.

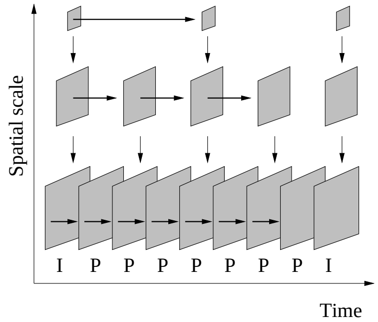

Figure 4: Spatio-temporal pyramid for scalable video coding. The original video sequence is represented at different spatial and temporal scales. Encoding is done by a pyramid codec as shown in Fig. 3 which can be easily extended to include motion compensated coding.

An example is shown in Fig. 4. [3] describes an efficient video codec based on a motion compensated spatio-temporal pyramid combined with E_8-lattice quantization which is used for scalable Internet video transmission. It combines an oversampled pyramid decomposition with motion compensation loops of the lowpass signals (LP_i of Fig. 3) in each layer.

Similar spatio-temporal resolution pyramids have recently also been included as scalable extensions of DCT-based motion-compensated video coding standards, such as ITU-T H.263+ or ISO MPEG-4.

4 CONCLUSION

Multiresolution image coding is based on the analysis of an image at a hierarchy of resolutions or scales. For compression, subband pyramid decompositions are a good compromise between frequency selectivity and spatial localization. Critically sampled subband decompositions have been successfully applied to image coding. We argue that, for applications requiring scalability, it is advantageous to use oversampled pyramid decompositions. Analysis and synthesis filters can be chosen freely and motion compensation can be easily incorporated for scalable video coding. An important problem in multiresolution coding is bit allocation. The well-known analytical solution for critically sampled Gaussian subband signals has recently been generalized to open-loop and closed-loop coding of oversampled pyramid decompositions. These solutions as well as simulations with natural images show that closed-loop coding has several advantages over open-loop coding which makes it more suitable for practical applications.

References

[1] B. Girod. Scalable video for multimedia systems. *Computers & Graphics*, 17(3):269–276, 1993.

[2] D. Taubmann and A. Zakhor. Multirate 3-D subband coding of video. *IEEE Trans. on Image Processing*, (5):572–588, Sep. 1994.

[3] U. Horn and B. Girod. Scalable video transmission for the internet. *Computer Network and ISDN Systems*, 29(15):1833–1842, Nov. 1997.

[4] M. Vetterli and J. Kovačević. *Wavelets and Subband Coding*. Signal Processing. Prentice-Hall, Englewood Cliffs, NJ, 1995.

[5] J.W. Woods (ed.). *Subband Image Coding*. Kluwer Academic Publishers, Boston, 1991.

[6] A. N. Akansu and R. A. Haddad. *Multiresolution signal decomposition*. Academic Press, San Diego, 1992.

[7] A. N. Akansu and M. J. T. (eds.) Smith. *Subband and Wavelet Transforms*. Kluwer Academic Publishers, Norwell, MA, 1995.

[8] M. Vetterli and K.M. Uz. Multiresolution coding techniques for digital television: A review. *Multidimensional Systems and Signal Processing*, 3:161–187, 1992.

[9] R. L. Lagendijk, F. Bosveld, and J. Biemond Subband video coding. In Ali N. Akansu and Mark J.T. Smith, editors, *Subband and Wavelet Transforms*, chapter 8, pages 251–285. Kluwer Academic Publishers, Norwell, MA, 1995.

[10] B. Girod, F. Hartung, and U. Horn. Subband image coding. In Ali N. Akansu and Mark J.T. Smith, editors, *Subband and Wavelet Transforms*, chapter 7, pages 213–250. Kluwer Academic Publishers, Norwell, MA, 1995.

[11] J. M. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41(12):3445–3462, Dec 1993. Special Issue on Wavelets and Signal Processing.

[12] A. Said and W. Pearlman. Image compression using the spatial-orientation tree. In *in Proc. IEEE Int. Symp. Circ. and Syst.*, pages 279–282, 1993.

[13] M.K. Uz, M. Vetterli, and D.J. LeGall. Interpolative multiresolution coding of advanced television with compatible subchannels. *IEEE Trans. on Circuits and Systems for Video Technology*, 1(1):86–99, Mar. 1991.

[14] P.J. Burt and E.H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.*, COM-31:532–540, Apr. 1983.

[15] B. Aiazzi, L. Alparone, S. Baronti, and F. Lotti. Lossless image compression by quantization feedback in a content-driven enhanced laplacian pyramid. *IEEE Trans. on Image Processing*, 6(6):831–843, Jun. 1997.

[16] T. Senoo and B. Girod. Vector quantization for entropy coding of image subbands. *IEEE Transactions on Image Processing*, 1(3), Oct. 1993.

[17] K. Ramchandran, A. Ortega, and M. Vetterli. Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders. *IEEE Trans. on Signal Processing*, 3(5):533–545, Sep. 1994.

[18] U. Horn, T. Wiegand, and B. Girod. Bit allocation methods for closed-loop coding of oversampled pyramid decompositions. In *Proc. ICIP'97*, volume II, pages 17–20, Santa Barbara, Oct. 1997.

[19] U. Horn and T. Wiegand. Calculations to solve the bit allocation problem for oversampled pyramid decompositions with and without noise feedback. URL: http://www-nt.e-technik.uni-erlangen.de/~wiegand/icip97-proof.ps.gz, Aug. 1997.

[20] L. Alparone. Quantization noise feedback in Laplacian pyramid-based image coding: A rate-distortion approach. In *Proc. IEEE DSP 97*, pages 849–852, Santorini, Jul. 1997.

[21] M. Pecot, P. J. Tourtier, and Y. Thomas. Compatible coding of television images, Part 1. Coding algorithm. *Signal Processing: Image Communication*, 2(3):245–258, Oct. 1990.

[22] M. Pecot, P. J. Tourtier, and Y. Thomas. Compatible coding of television images, Part 2. Compatible systems. *Signal Processing: Image Communication*, 2(3):259–268, Oct. 1990.

[23] P. J. Tourtier, M. Pecot, and J.-F. Vial. Motion compensated subband coding schemes for compatible high definition TV coding. *Signal Processing: Image Communication*, 4:325–344, 1992.

[24] J.-R. Ohm. Three-dimensional subband coding with motion compensation. *IEEE Trans. on Image Processing*, 3(5):559–571, Sep. 1994.

[25] T. Naveen and J. W. Woods. Motion compensated multiresolution transmission of high definition video. *IEEE Trans. on Circuits and Systems for Video Technology*, 4(1):29–41, Feb. 1994.